DOCUMENT RESUME

ED 099 396 TM 004 020

AUTHOR Martuza, Victor R.; Engel, John D.
TITLE The Effect of Substituting p for alpha on the Unconditional and Conditional Powers of a Null Hypothesis Test.
PUB DATE [Apr 74]
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Bayesian Statistics; *Hypothesis Testing; *Statistical Analysis; Validity
IDENTIFIERS *Power (Statistics)

ABSTRACT

Results from classical power analysis (Brewer, 1972) suggest that a researcher should not set a=p (when p is less than a) in a posteriori fashion when a study yields statistically significant results because of a resulting decrease in power. The purpose of the present report is to use Bayesian theory in examining the validity of this generalization. Using the t-test and Bayes' theorem we show that while the substitution of p for a (when p is less than a) reduces unconditional or simple power, it actually increases the conditional power of the test--i.e., the probability that the alternative hypothesis is true given a statistically significant result. Because of this, the substitution of p for a a posteriori (when p is less than a), i.e., acting as though the value of a is equal to the obtained p value, seems to be a harmless practice at worst, and on the positive side, provides important information to the research consumer interested in the validity of the alternative hypothesis given the data of the study. (Author/RC)

ED 099 96

The Effect of Substituting p for alpha On The Unconditional and Conditional Powers of A Null Hypothesis Test

Abstract

Results from classical power analysis (Brewer, 1972) suggest that a researcher should not set $\alpha = p$ (when $p < \alpha$) in an *a posteriori* fashion when a study yields statistically significant results because of a resulting decrease in power. The purpose of the present report is to use Bayesian theory in examining the validity of this generalization. Using the *t* test and Bayes' theorem we show that while the substitution of p for α (when $p < \alpha$) reduces unconditional or simple power, it actually increases the conditional power of the test--i.e., the probability that H_1 is true given a statistically significant result. Because of this, the substitution of p for α *a posteriori* (when $p < \alpha$), i.e., acting as though the value of α is equal to the obtained p value, seems to be a harmless practice at worst, and on the positive side, provides important information to the research consumer interested in the validity of the alternative hypothesis (H_1) given the data of the study.

The Effect of Substituting p for alpha On The Unconditional and Conditional Powers of a Null Hypothesis Test

February 24, 1974

Recently, a good deal of attention has been focused on the area of power analysis in the behavioral sciences. Cohen (1969), Overall (1969), Tversky and Kahnman (1971), and Brewer (1972, 1973) among others, have strongly suggested that the power (the probability of rejecting the null hypothesis (H_o) when the alternative hypothesis (H_1) is true) of a statistical test should be determined prior to conducting an experiment. The importance of this analysis is suggested by Tversky and Kahnman (1971): (1)the results of a power analysis may indicate that the study should not be conducted unless the sample size is substantially increased; (2)power analysis is necessary to the explanation of negative results; (3)the results of power analysis provide an indication of the probability of correctly rejecting H_o.

Cohen's concern with power analysis in the behavioral sciences has resulted in the first easily manageable set of procedures and tables for the calculation of the power of statistical tests. The use of Cohen's tables to determine the power of a test requires the researcher to combine a number of readily available pieces of information: (1)sample size, (2) α level, and (3)effect size (ES) which Cohen has defined as "the degree to which the phenomenon is present in the population."

As part of this trend, Brewer (1973) has conducted an extensive analysis of the statistical tests appearing in several educational and psychological research journals during the last few years and found that the power in many studies was inadequate. Brewer (1972) has discussed a number of procedures for maximizing the power of statistical tests used in behavioral science research. One of these is simply to avoid the "practice of taking a sample, calculating p (the probability of the statistic under H_o) and then equating this value to α (when $p < \alpha$)." He contends that the minimum sample size necessary to detect a significant departure from the null condition at $\alpha = .05$ is altogether too small to detect the same departure at $\alpha = .01$ because change in the size of α is directly related to change in the power of the test. He provides the following as an example:

> If a t test is conducted with $n_1 = n_2 = 50$ and medium ES, then power is approximately .70 for $\alpha = .05$ and .45 for $\alpha = .01$ ... It is clear that if power is to be maintained at .70 for this sample and ES, then the test is not significant at the .01 level even if $p \leq .01$. The fact that $p \leq .01$ only allows the researchers to reject H_o at the $\alpha = .05$ level under these conditions. (Researchers may, of course, equate p and α if they are willing to take the resulting reduction in power.) (p. 395)

Here, Brewer argues against substituting p for α (when $p < \alpha$) strictly because the adoption of such a decision rule reduces the power of the test. While others (e.g., Cohen, 1973; Dayton, et al, 1973) have pointed out that the power argument advanced by Brewer is inappropriate within the classical (non-Bayesian) framework, a demonstration of the effects of such a policy on the conditional power of a null hypothesis test seems useful at this time. In the following discussion, Bayes' theorem and the t test are used for this purpose.

A Bayesian Perspective

Overall (1969) has noted that particular concepts from Bayesian theory have logical implications for the design of experiments. Bayes theorem provides a formal statement of the probability of the null hypothesis being true given a statistically significant result, and may be stated as:

$$P(H_o \mid D) = \frac{P(D \mid H_o) \cdot P(H_o)}{P(D)}$$

where $P(H_o \mid D)$ is the probability of the H_o being true given a statistically significant result (D); $P(D \mid H_o)$ is the probability of significant result given a true null hypothesis; $P(H_o)$ is the a priori probability that the null hypothesis is true and P(D) is the probability of a statistically significant result when either the null or alternative hypothesis is true. Overall defines a conditional alpha probability, $P(H_o \mid D)$, as "the probability that a result which is judged statistically significant has occurred by chance rejection of a valid null hypothesis." He illustrates this difference between the simple and conditional α levels by means of the following example:

> Consider the case of an investigator who works entirely within the area where the null hypothesis is always true. If he consistently employs the proper classical statistical model with $\alpha = .05$ or $\alpha = .01$, the probability will be only .05 or .01 that he will reject a null hypothesis when it is in fact valid; however, of the statistically significant results reported by this hapless investigator, 100% will be due to chance. This ad absurdum example illustrates the difference between simple alpha probabilities and conditional alpha probabilities... (p. 286)

What is the practical significance of conditional alpha probability? To answer this question, Overall cites a case which he claims exemplifies psychiatric drug research. Typically, this area is characterized by small sample size, low power tests, and unlikely treatment differences. Given these conditions and the simple α level of .05, the conditional alpha level was calculated to be .643. Overall concludes that the analysis indicates that approximately two-thirds of the statistically significant results will be type I errors!

In light of Overall's analysis, an examination of the ramifications of substituting p for α seems in order. For while substituting p for α (when $p < \alpha$) clearly reduces unconditional or simple power,[1] this particular notion of power is irrelevant once the data have been inspected. However, the effect of this action on conditional power (i.e., the probability that H_1 is true given a statistically significant result) seems appropriate and is the focus of the remainder of this discussion.

In terms of Bayes' theorem, conditional power is defined as:

$$P(H_1|D) = \frac{P(D|H_1)\ P(H_1)}{P(D)}$$

where $P(H_1|D)$ is the probability of a true alternative hypothesis given a statistically significant result (D); $P(H_1)$ is the <u>a priori</u> probability that the alternative hypothesis is true; $P(D|H_1)$ is the probability of a significant result (D) given that H_1 is true; P(D) is the probability of a statistically significant result when either H_o or H_1 is true--i.e., $P(D|H_1)\ P(H_1) + P(D|H_o)\ P(H_o)$.

We are now in the position to demonstrate the effects of substituting p for α (when $p < \alpha$) on the conditional and unconditional power of a statistical test. The variables of interest are:

a. <u>Effect Size (ES)</u> - the degree to which the phenomenon exists (Cohen, 1969). Cohen's small (.20), medium (.50) and large (.80) ES were used.

b. <u>Sample Size</u> - small (n = 20), medium (n = 50) and large (n = 100) sample sizes were used.

c. <u>$P(H_o)$</u> - the <u>a priori</u> probability of the null hypothesis <u>being</u> true: small (.20), medium (.50), and large (.90) prior probabilities were used.

[1] It is recognized that the term "unconditional or simple" power is not entirely appropriate since power is always conditional on a number of factors. However, the term as used here refers to power in the classical sense.

d. Alpha Level - .05 was selected.

e. p-level - .01 was selected.

f. Statistical Test - Student's *t* test was selected.

For each combination of effect size, sample size, $P(H_o)$, alpha level and p-level, the conditional alpha probability $P(H_o|D)$ and the conditional power probability $P(H_1|D)$ were calculated. The results of these calculations are shown in Table 1. Notice that for a given ES, sample size, and $P(H_o)$, the unconditional or simple power probability of the *t* test decreases when simple α is reduced from .05 to .01 while the conditional power probability of the test actually increases.

While attention is still focused on the table, two other relationships may be noted:

a. As the effect size (ES) increases, the probability of a valid null hypothesis given a significant result $[P(H_o|D)]$ decreases at each level of sample size and each level of the probability of a valid H_o $[P(H_o)]$ whereas the probability of a valid alternative hypothesis given a significant result $[P(H_1|D)]$ increases at each level of sample size and $P(H_o)$.

b. As sample size increases, the probability of a null hypothesis given a significant result $[P(H_o|D)]$ decreases at each level of effect size (ES) and each level of the probability of a valid H_o $[P(H_o)]$ whereas the probability of a valid alternative hypothesis given the data $[P(H_1|D)]$ increases at each level of ES and $P(H_o)$.

Conclusions

It is clear that if p is substituted for α (when $p < \alpha$) *a posteriori*, then power in the classical sense is reduced. But, clearly, this unconditional power is no longer relevant once the data have been inspected. Following the notions of Overall and others (e.g., Rozeboom, 1960), one must conclude that if power must be calculated, only conditional power is relevant.

From the Bayesian perspective, substituting p for α (when $p < \alpha$) has no adverse effect on the measure of power and, on the positive side, provides the consumer who is interested in the tenability of H_1 given the data with useful information.

References

Brewer, J. K. On the power of statistical tests. American Educational Research Journal, 1972, 9(3), 391-401.

Brewer, J. K. & Owen, P. W. A note on the power of statistical tests. Journal of Educational Measurement, 1973, 10(1), 71-74.

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.

Cohen, J. Statistical power analysis and research results. American Educational Research Journal, 1973, 10(3), 225-23C.

Dayton, C. M., Schafer, W. D. & Rogers, B. C. On appropriate uses and interpretations of power analysis: A comment. American Educational Research Journal, 1973, 10(3), 231-234.

Overall, J. E. Classical statistical hypothesis testing within the context of Bayesian theory. Psychological Bulletin, 1969, 71, 285-292.

Rozeboom, W. W. The fallacy of the null-hypothesis significance test. Psychological Bulletin, 1960, 57, 416-428.

Tversky, A. & Kahnman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.

Table 1

Conditional alpha probability [$P(H_o|D)$] and conditional pow

for selected values of effect size, sample size, a priori prob

being true [$P(H_o)$], alpha level and p

		Small Effect Size (.20)									Medium Effect Size (..						
		Small Sample Size (20)			Medium Sample Size (50			Large Sample Size (100)			Small Sample Size (20)			Medium Sample Size (50)			
		$P(H_o)$			$P(H_o)$			$P(H_o)$			$P(H_o)$			$P(H_o)$			
		Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	
Unconditional $\alpha = .05$	$P(H_o\|D)$	.12	.36	.83	.07	.23	.73	.04	.15	.61	.04	.13	.58	.02	.07	.39	
	$P(H_1\|D)$	.88	.64	.17	.93	.77	.27	.96	.85	.39	.96	.87	.42	.98	.93	.61	
	$P(D\|H_1)$	.09	.09	.09	.17	.17	.17	.29	.29	.29	.33	.33	.33	.70	.70	.70	
Unconditional $p = .01$	$P(H_o\|D)$	.11	.33	.82	.04	.14	.60	.02	.08	.43	.02	.07	.39	.00	.02	.17	
	$P(H_1\|D)$	.89	.67	.18	.96	.86	.40	.98	.92	.57	.98	.93	.61	1.00	.98	.83	1
	$P(D\|H_1)$	.02	.02	.02	.06	.06	.06	.12	.12	.12	.14	.14	.14	.45	.45	.45	

Table 1

ility [$P(H_o|D)$] and conditional power probability [$P(H_1|D)$]

ct size, sample size, *a priori* probability of the null hypothesis

ing true [$P(H_o)$], alpha level and p - level

Medium Effect Size (.50)									Large Effect Size (.80)								
Small Sample Size (20)			Medium Sample Size (50)			Large Sample Size (100)			Small Sample Size (20)			Medium Sample Size (50)			Large Sample Size (100)		
$P(H_o)$			$P(H_o)$			$P(H_o)$			$P(H_o)$			$P(H_o)$			$P(H_o)$		
Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)	Small (.20)	Medium (.50)	Large (.90)
.04	.13	.58	.02	.07	.39	.01	.05	.32	.02	.07	.39	.01	.05	.31	.01	.05	.31
.96	.87	.42	.98	.93	.61	.99	.95	.68	.98	.93	.61	.99	.85	.69	.99	.95	.69
.33	.33	.33	.70	.70	.70	.94	.94	.94	.69	.69	.69	.98	.98	.98	1.00	1.00	1.00
.02	.07	.39	.00	.02	.17	.00	.01	.10	.00	.02	.17	.00	.01	.09	.00	.00	.08
.98	.93	.61	1.00	.98	.83	1.00	.99	.90	1.00	.98	.83	1.00	.99	.91	1.00	1.00	.92
.14	.14	.14	.45	.45	.45	.82	.82	.82	.44	.44	.44	.91	.91	.91	1.00	1.00	1.00